

ECIR 2017 Workshop on Exploitation of Social Media for Emergency Relief and Preparedness (SMERP 2017)

Saptarshi Ghosh

Indian Institute of Technology, Kharagpur, India
Indian Institute of Engineering Science and Technology, Shibpur, India
saptarshi@cse.iitkgp.ernet.in, sghosh@cs.iiests.ac.in

Kripabandhu Ghosh

Indian Institute of Technology, Kanpur, India
kripa@cse.iitk.ac.in, kripa.ghosh@gmail.com

Debasis Ganguly

IBM Research Labs, Dublin, Ireland
debasis.ganguly1@ie.ibm.com, debforit@gmail.com

Tanmoy Chakraborty

University of Maryland, College Park, USA
tanchak@umiacs.umd.edu

Gareth J.F. Jones

Dublin City University, Ireland
Gareth.Jones@computing.dcu.ie

Marie-Francine Moens

Katholieke Universiteit Leuven, Belgium
marie-francine.moens@cs.kuleuven.be

Abstract

The first international workshop on Exploitation of Social Media for Emergency Relief and Preparedness (SMERP) was held in conjunction with the 2017 European Conference on Information Retrieval (ECIR) in Aberdeen, Scotland, UK. The aim of the workshop was to explore various technologies for extracting useful information from social media content in disaster situations. The workshop included a peer-reviewed research paper track, a data challenge, two keynote talks, and discussion sessions on the relevant open research challenges. This report presents an overview of the workshop, including the motivations behind organizing the workshop, and summaries of the research papers and keynote talks at the workshop. We also reflect on the future directions as inferred from discussion sessions during the workshop.

1 Introduction and Motivation

A huge volume of user-generated content in online social media (OSM) platforms like Twitter, Facebook, WhatsApp, etc. can potentially become a real-time and reliable source of information for mitigating disaster situations. In particular, during emergency situations (e.g., natural disasters such as earthquakes, cyclones, floods, and man-made disasters such as terror attacks, riots), information posted on OSM can contribute significantly to relief operations [2, 4]. Moreover, content from OSM may also be used for emergency preparedness, such as for developing early warning systems.

The primary objective behind organising the SMERP workshop¹ was to provide a platform for researchers to present their work on utilising social media content for improving emergency relief operations and emergency preparedness initiatives. In addition to a peer-reviewed research paper track, the workshop included a data challenge track. In this data challenge track, a set of microblogs posted during a recent disaster event was shared, and participating teams developed methodologies for two tasks – *Text retrieval* and *Text summarization*. Details of these two tracks are presented in Sections 1.1 and 1.2.

1.1 Peer-review Track

The workshop invited original research contributions related to the theme, which included the following topics:

- Information retrieval and extraction from short, noisy content posted on OSM.
- Applications of data mining, NLP and machine learning for processing OSM content.
- Aggregating information from multiple OSM and online / offline resources.
- Addressing the code-mixed and informal vocabulary of OSM content.
- Detection of events and emerging themes.
- Real-time management and summarization of dynamic content streams.
- Detection of rumours, and identification of trustworthy sources and information.
- Geo-tagging and geo-localisation of content and sources.
- Social network models for information diffusion in emergency situations.
- Identifying disaster-prone or accident-prone regions and infrastructures.
- Crowdsourcing systems for emergency preparedness and disaster relief.
- Mining interactions among emergency preparedness and relief groups.

1.2 Data Challenge Track

The workshop endeavoured to promote development of IR methodologies for some practical challenges that need to be addressed during an emergency event, along with thorough evaluation and comparison of the methodologies. To this end, a data challenge was organised, following the style of TREC tracks [3]. A dataset of 72,220 tweets posted during a recent disaster event – the earthquake in Italy in August 2016 – was provided, along with a set of specific information needs (topics) during an emergency event, such as what resources are needed, what resources are available, what are the casualties and infrastructure damages, and so on. Participants were invited to submit solutions to the two tasks, *Text Retrieval* task and *Text Summarization* task, described in Sections 1.2.1 and 1.2.2 respectively.

1.2.1 Text Retrieval

In this task, the participants were asked to develop methodologies for extracting tweets that are relevant to the given set of topics. This problem had two levels – Level 1 and Level 2. In Level 1, 52,469 tweets posted during the first day (first 24 hours) after the earthquake were provided.² The participants were then asked to extract tweets relevant to each specified topic. Relevance assessment was done by human annotators on the set of tweets submitted by the participants. In Level 2, 19,751 tweets posted during the second and third days after the earthquake (i.e., during the 48-hour interval after the first 24 hours after the earthquake) were provided. In addition, the relevance assessments of the tweets submitted by the participants in Level 1, was also provided. The participants were allowed to use these assessments as training for their Level 2 submissions.

¹Workshop website: <http://computing.dcu.ie/~dganguly/smerp2017/>

²Following the terms and conditions of Twitter, only the tweet-ids were provided, along with a script to download the tweets using the Twitter API.

1.2.2 Text Summarization

In this task, the participants were asked to develop summarization techniques on the tweets that are relevant to each topic. Similar to the Text Retrieval task, this task also had two levels to serve the same purpose. That is, in Level 1 the participants were asked to submit summaries of the tweets posted during the first day (24 hours) after the earthquake, on each of the topics. These summaries were evaluated by comparing with gold standard summaries generated by human annotators. In Level 2, the tweets collected during the second and third days after the earthquake was given to the participants for summarization. Additionally, an assessment of the summaries submitted by the participants in Level 1 was also provided – the parts of the summaries that were common with the gold standard summaries were indicated. The participants were allowed to use the assessments to improve their summarization algorithm for Level 2.

In both tasks, the idea behind having two levels was that the participants could improve their algorithms using the relevance assessments in Level 1, so that the performance could be improved in Level 2. In practice, this setup was an attempt to capture the effects of temporal variation in social media streams, where algorithms need to be tuned to capture the temporal variations.

For both the tasks, we allowed three types of submissions:

- *Full-Automatic* – No manual intervention at any stage
- *Semi-automatic* – manual intervention involved in query formulation stage (but not in the retrieval / summarization stage)
- *Manual* – manual intervention involved in both query formulation and retrieval / summarization stages

2 Activities at the Workshop

We now describe the activities at the workshop – the keynote talks, the papers accepted through the peer-reviewed track and presented at the workshop, and the results of the data challenge.

2.1 Keynote Talks

The following two keynote talks were delivered. They gave a good view of the state-of-the-art to the participants. Also, the keynote speakers touched upon some future research directions.

- *Title:* Time-Critical Analysis of Evolving Social Media Streams During Sudden-Onset Events
Speaker: Muhammad Imran, Qatar Computing Research Institute, Qatar
Abstract: Sudden-onset emergencies such as natural or man-made disasters bring uncertainties in which time-critical information needs emerge from formal response organizations, affected communities and other concerned population. The growing adaption of Information and Communication Technologies (ICT) and Social Networks such as Twitter, Facebook has created numerous opportunities to disseminate and consume critical information during an on-going situation. However, time-critical analysis of high-velocity social media streams containing high-volume data involves solving multiple challenges including real-time parsing of brief and informal messages, handling information overload issues, and classifying, summarizing, and prioritizing different types of information.
- *Title:* Summarizing the Situation with Social Media Streams
Speaker: Richard McCreadie, School of Computing Science, University of Glasgow, UK
Abstract: When a crisis hits, it is important for response agencies to quickly determine the situation on the ground, such that they can deploy the limited resources at their disposal as quickly and effectively as possible. However, during an emergency, information is difficult to come-by, as response units often need to arrive on the scene before the severity of the situation

can be estimated. On the other hand, during emergencies, the general public is gravitating to social media platforms to ask for assistance and to show what they see to their friends. As such, emergency services are increasingly interested in technologies that can extract relevant information from social media during an emergency, to aid situational awareness. Meanwhile, real-time summarization is an emerging field that aims to build timeline summaries of events that are happening in the world, using news and social media streams as sensors. In this talk, I will provide an overview of what types of information emergency services want to extract from social media, and how real-time summarization systems can help achieve this. Furthermore, I will discuss current technologies and techniques for real-time summarization that are relevant to the crisis domain, along with the challenges that are yet to be solved [1].

2.2 Peer-review Track

Five papers were accepted and presented in this track. The papers are described as follows.

- *Title:* Weakly Supervised Classification of Tweets for Disaster Management.
Authors: Girish Keshav Palshikar, Manoj Apte and Deepak Pandita.
Abstract: Social media has quickly established itself as an important means that people, NGOs and governments use to spread information during natural or man-made disasters, mass emergencies and crisis situations. Given this important role, real-time analysis of social media contents to locate, organize and use valuable information for disaster management is crucial. In this paper, we propose self-learning algorithms that, with minimal supervision, construct a simple bag-of-words model of information expressed in the news about various natural disasters. Such a model is human-understandable, human-modifiable and usable in a real-time scenario. Since tweets are a different category of documents than news, we next propose a model transfer algorithm, which essentially refines the model learned from news by analyzing a large unlabeled corpus of tweets. We show empirically that model transfer improves the predictive accuracy of the model. We demonstrate empirically that our model learning algorithm is better than several state of the art semi-supervised learning algorithms.
- *Title:* Identification of disaster-implicated named entities.
Authors: Dávid Márk Nemeskey, Ivett Benyeda and András Kornai.
Abstract: For disaster preparedness, a key aspect of the work is the identification, ahead of time, of potentially implicated locations (LOC), organizations (ORG), and persons (PER). Here we describe how static repositories of traditional news reports can be rapidly exploited to yield disaster- or accident-implicated named entities.
- *Title:* Microblog Retrieval in a Disaster Situation: A New Test Collection for Evaluation.
Authors: Moumita Basu, Anurag Roy, Kripabandhu Ghosh, Somprakash Bandyopadhyay and Saptarshi Ghosh.
Abstract: Microblogging sites are important sources of situational information during disaster situations. Hence it is important to design and evaluate Information Retrieval (IR) systems that retrieve information from microblogs during disaster situations. The primary contribution of this paper is to develop a test collection for evaluating IR systems for microblog retrieval in disaster situations. The collection consists of about 50,000 microblogs posted during the Nepal earth-quake in April 2015, a set of five topics (information needs) that are practically important during a disaster, and the gold standard annotations of which microblogs are relevant to each topic. We also present some IR models that can be suitable in this evaluation setup, including a standard language model based retrieval, and word embedding based retrieval. We find that the term embedding based retrieval performs better for short, noisy microblogs.
- *Title:* Situational Awareness for Low Resource Languages: the LORELEI Situation Frame Annotation Task.
Authors: Stephanie Strassel, Ann Bies and Jennifer Tracey.
Abstract: The objective of the LORELEI Situation Frame task is to aggregate information from multiple data streams - including social media - into a comprehensive, actionable understanding

Team-id	Member names	Affiliation	Participation
CSPIT	Rutvika Nanacha, Sindur Patel	C.S.Patel Institutes of Technology, and Charusat University, Changa, India	Both tasks
DAICT	Sandip Modha, Rishikesh Bhatt, Chintak Mandalya	IR Lab, Dhirubhai Ambani Institute of Information and Communication Tech- nology, Gandhinagar, India	Both tasks
DCU	Wei Li	School of Computing, Dublin City Uni- versity, Dublin, Ireland	Information Extraction only
IEST	Kanav Mehra, Vibhash Chandra	Indian Institute of Engineering Science and Technology, Shibpur, India	Summarization only
RU	Hürriyetoglu, A. (Ali), Nelleke Oostdijk	Centre for Language Studies, Radboud University, Nijmegen, The Netherlands	Information Extraction only
USI	Anastasia Giachanou, Ida Mele, Fabio Crestani	Universita della Svizzera Italiana (USI), Lugano, Switzerland	Both tasks

Table 1: Teams participating in the data challenge track (listed in the alphabetic order of team-id).

of the basic facts needed to mount a response to an emerging situation. Rather than evaluating these capabilities in English, LORELEI is particularly concerned with advancing human language technology performance for low resource languages. The combination of domain, genre and language requirements make creation of linguistic resources for LORELEI in general, and the Situation Frame task in particular, especially challenging. Data is by definition relatively scarce for these languages, and real operational data may be impossible to come by, necessitating the use of “proxy” data sources. The annotation task itself, while superficially straightforward, requires navigating many difficult decisions involving the use of inference and the presence of widespread ambiguity and under-specification in the source data. We introduce the Situation Frame annotation task in the context of the goals of the larger LORELEI program, explore some of the most prevalent annotation challenges, and discuss the impact of various data types on annotation consistency. The data described in this paper will be made available to the wider research community after its use in LORELEI program evaluations.

- *Title:* Microblog Retrieval During Disasters: How To Create Ground Truths?
Authors: Ribhav Soni and Sukomal Pal
Abstract: Microblogging services like Twitter are an important source of real-time information during disasters and can be utilized to aid rescue, relief and rehabilitation efforts. The focus of this work is on the creation of gold standard data for automatic retrieval of helpful tweets. Using various experiments on the gold standard data prepared in the FIRE 2016 Microblog Track, we show that the gold standard data prepared missed many relevant tweets. We also demonstrate that using a machine learning model can help in retrieving the remaining relevant tweets by training an SVM model on a subset of the data and using it to get the most useful tweets in the entire dataset. We obtain high precision and recall even with very little training data, which makes such a model suitable for use in a real-time disaster situation.

Full versions of the above papers can be accessed at the workshop website <http://computing.dcu.ie/~dganguly/smerp2017/>.

2.3 Data Challenge Track

A total of six teams participated in the data challenge track. The teams (team-ids, team member names and their affiliations) along with their participation details are shown in Table 1.

2.3.1 Text Retrieval Task

The primary evaluation measure was *Bpref*. In addition, Precision@20, Recall@1000 and Mean Average Precision (MAP) values were also reported.

Only team DCU submitted *Full-automatic* runs. In Level 1, their best Bpref value was 0.6170 and in Level 2 their best Bpref value was 0.7767. Among the *Semi-automatic* submissions in Level 1, DCU produced the best performance with a Bpref value of 0.4407. In Level 2, among the *Semi-automatic* submissions, team RU produced the best performance with a Bpref value of 0.4724.

2.3.2 Text Summarization Task

The primary evaluation measure was ROUGE-L. In addition, ROUGE-1, ROUGE-2 and ROUGE-SU4 values were also reported.

In Level 1, only team USI submitted *Full-automatic* summaries and their best performance in terms of ROUGE-L score was 0.3029. In Level 1, the best *Semi-automatic* summary was submitted by team IEST with a ROUGE-L score of 0.4885. In Level 2, no team submitted *Full-automatic* summaries. In this level also, the best performance was produced by the IEST team with a ROUGE-L score of 0.5142.

3 Discussion Sessions

The workshop featured multiple discussion sessions, where the participants deliberated on the open research questions and challenges in the area of utilising social media for disaster informatics. Participants generally agreed that, although there exists a considerable amount of research work in this area, there are still several unsolved challenges.

For instance, organisations participating in post-disaster relief operations are usually consulted to get an idea of the types of information that need to be extracted from social media content. However, often the organisation members are not fully aware of the widely different types of information available on social media; hence new categories of information might need to be added and extracted. Also, the categories of information should be changed dynamically with time during an ongoing disaster event, as new information emerges.

Again, multi-lingual and code-mixed content are becoming more and more common on social media, especially, where there are disasters in developing countries, a lot of valuable information is posted in local resource-poor languages. Effective semantic matching techniques need to be developed to utilise such content. Another direction that remains relatively unexplored is to combine information from multiple modes or multiple sources, such as text, image, sensor data, and so on.

Lastly, there was also discussion on designing practical data challenges and developing benchmark data collections on which IR techniques can be evaluated and compared.

4 Acknowledgements

The organisers of SMERP acknowledge the support of several persons without whose help the workshop could not be organised. The peer-review track would not be possible without the support of the Program Committee members – the full list is available on the workshop website. For the data challenge track, the support of the human annotators (who developed the gold standard for evaluation) is acknowledged, specially that of Moumita Basu and Somenath Das. Also acknowledged is partial support by a grant from the Information Technology Research Academy (ITRA), DeITY, Government of India (Ref. No.: ITRA/15 (58)/Mobile/DISARM/05).

References

- [1] Matthew Ekstrand-Abueg, Richard McCreadie, Virgil Pavlu, and Fernando Diaz. A study of realtime summarization metrics. In *CIKM 2016*, pages 2125–2130, New York, NY, USA, 2016. ACM.
- [2] Muhammad Imran, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. Processing Social Media Messages in Mass Emergency: A Survey. *ACM Computing Surveys*, 47(4):67:1–67:38, June 2015.
- [3] Iadh Ounis, Craig Macdonald, Jimmy Lin, and Ian Soboroff. Overview of the TREC-2011 Microblog Track. Available at: <http://trec.nist.gov/pubs/trec20/papers/MICROBLOG.OVERVIEW.pdf>, 2011.
- [4] Sarah Vieweg, Amanda L. Hughes, Kate Starbird, and Leysia Palen. Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In *Proc. ACM SIGCHI*, 2010.